A guide to programme evaluation in quality improvement

What do we cover in this guide?

We describe an introductory six-step approach to planning and designing programme evaluation in quality improvement developed for use in the Improvement Hub. This approach outlines the range of evaluation questions that we consider and methods for answering these in the context of quality improvement.

For more detailed guidance about individual methods and tools we have included links at each step including those contained within the Rainbow Framework from Better Evaluation.
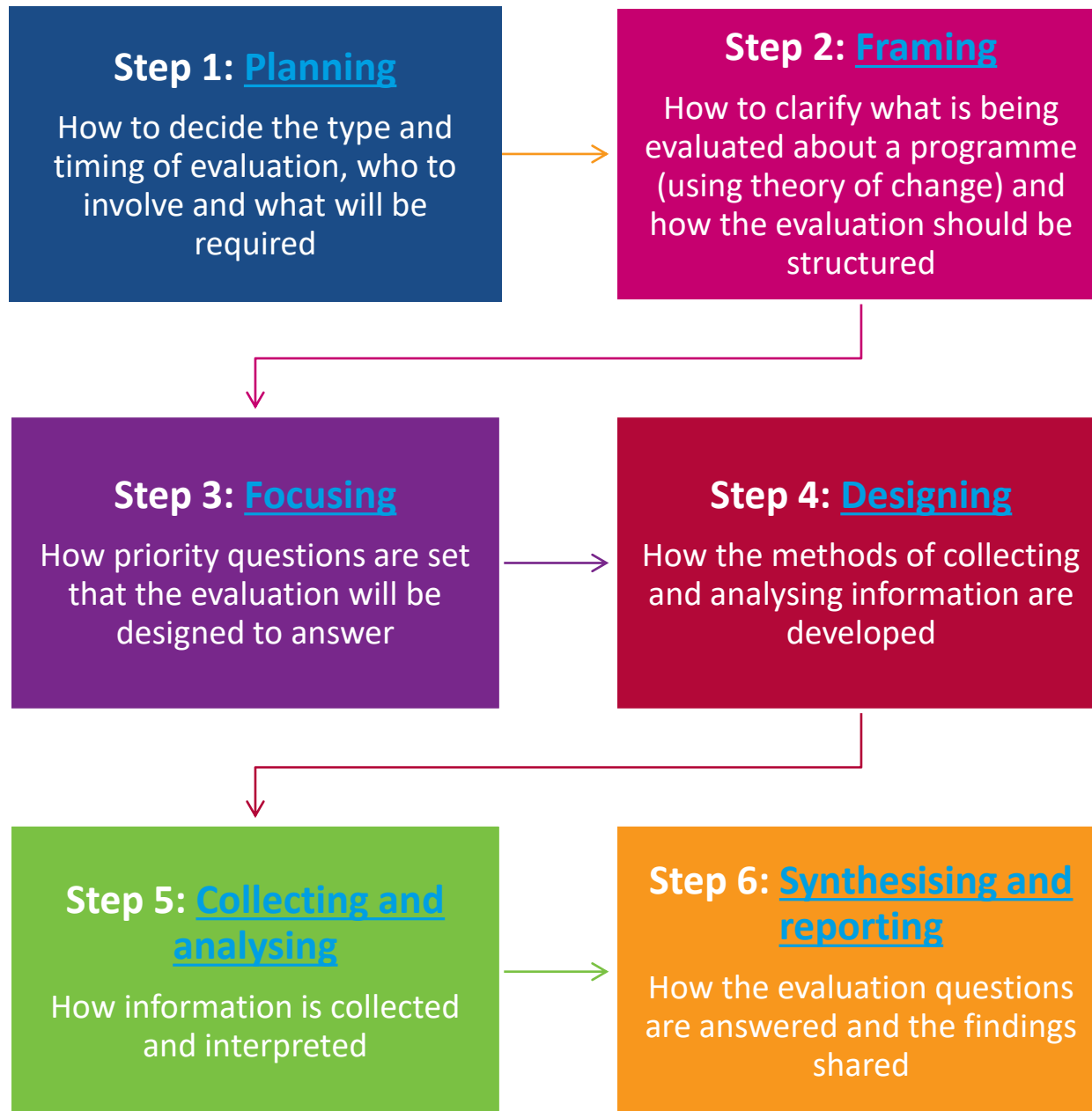
How do we define programme evaluation?

Programme evaluation allows us to better understand and capture the value of quality improvement work and inform decisions about scale and spread. It involves a process of collecting, analysing and synthesising information to answer priority questions about a programme according to different evaluation criteria (appropriateness, equity, process/implementation, efficiency, impact and sustainability).

A priority evaluation question might focus on impact – to what extent has a programme worked as expected? A broader and deeper question would consider whether and how a programme is the most impactful and equitable thing to do. Answering these different questions requires an understanding of how evaluation is planned and when and how different evaluation methods should be best used[1].

1. Parry GJ, Carson-Stevens A, Luff DF, McPherson ME, Goldmann DA. Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr*. 2013;13(6 Suppl):S23-S30.

Six-step approach

Step 1: Planning

How to decide the type and timing of evaluation, who to involve and what will be required

Step 2: Framing

How to clarify what is being evaluated about a programme (using theory of change) and how the evaluation should be structured

Step 3: Focusing

How priority questions are set that the evaluation will be designed to answer

Step 4: Designing

How the methods of collecting and analysing information are developed

Step 5: Collecting and analysing

How information is collected and interpreted

Step 6: Synthesising and reporting

How the evaluation questions are answered and the findings shared

Step 1: Planning: deciding when best to evaluate, how and who to involve

Planning sets out the purpose and scope of evaluation and starts to identify what will be required to achieve this. Making an assessment of the evaluability of a programme can support an effective plan to be developed.

Evaluability relates to whether and how a programme is likely to be evaluated in a way that will be useful and reliable.

There are a number different aspects of evaluability to consider:
- whether the programme can be described clearly enough including a theory of change to provide a basis for evaluation
- whether impact is plausible and measurable as a focus for evaluation
- what resource, expertise and data is likely to be available within the timeframe for evaluation that would influence the feasibility of using a particular evaluation approach

Key Points
- Consider the 'evaluability' of your programme when starting to plan – how will the programme design and availability of data affect how evaluation is conducted?
- Decide who will be managing the evaluation and when stakeholders will need to be involved

Alternatives to impact evaluation should be considered when decisions have been already been made about impact using other evidence or where data or resources are inadequate to support this type of evaluation.

For a programme in earlier stages of development and implementation, capturing how well the programme is being implemented and how this could be improved on (using real-time/formative methods) would be appropriate. Impact evaluation could then be designed to take place at the end of the programme.

In addition to evaluability, planning should also clarify how the process will be managed and how and when stakeholders will be involved in making decisions about the evaluation (a reference or steering group may be appropriate). A planning checklist can help to identify the key tasks and decisions before evaluation starts.

Step 1: Planning: deciding when best to evaluate, how and who to involve

What tools and resources can help?

There are a range of resources available from Better Evaluation about how to plan and manage an evaluation.

Evaluability assessment checklists can help structure and clarify the process of planning when and how evaluation should be carried out.

A step by step guide for engaging stakeholders in planning is available from Robert Wood Johnson Foundation.

Step 2: Framing: what is being evaluated about a programme (using theory)

Once there is plan in place, the next step is to develop a framework for the evaluation. This will set out the overall approach or strategy for the evaluation and the programme theory that this will be based on.

The programme theory articulates what is expected to improve over time (short, medium and long term outcomes), across what levels (personal, organisation, system), as a result of the successful implementation and delivery. How change will take place, where and for whom should also be differentiated.

This is sometimes referred to as a combination of execution theory and change theory which can be articulated together in a logic model. See more guidance on programme theory here.

Whether the needs of different groups are met for equitable outcomes is an important focus. Whether equity is primary concern for a QI programme or a more general one evaluation should consider how differential needs and outcomes will be assessed.

Key Points

- Think through how change will occur from the activities that are planned that will provide the focus for evaluating whether this happened (impact) and how (implementation/process)
- Establish detail around your outcomes with regards to what will be meaningful and realistic to evaluate
- Consider how differential outcomes are being understood

As well as defining what is known or expected about the programme, it also important for evaluation to consider what will be emergent and uncertain.

In complex systems change many different factors including adoption in different contexts and varying needs across the population influence success. This requires a broader and less linear understanding of change.

Structuring an evaluation to pay attention to the multiple factors influencing success can be supported using existing theoretical frameworks. See more guidance here.

Step 2: Framing: what is being evaluated about a programme (using theory)

What tools and resources can help?

There are a number of QI tools that can help with clarifying and developing a programme theory including those in the [QI essentials toolkit](#).
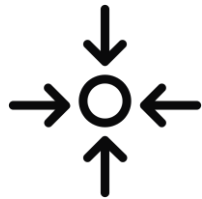
There are many resources that can help with defining the theory of change for a programme and what outcomes are meaningful including the [personal outcomes approach](#).

Step 3: Focusing: the questions the evaluation will be designed to answer

The next step is to focus the evaluation in more detail by agreeing the questions that the evaluation will be designed to answer. There are a range of overarching or key questions that could be prioritised across different evaluation criteria (appropriateness, equity, process, impact, efficiency and sustainability).

For an evaluation to be successful, it is important to prioritise these questions based on a shared understanding of the evaluation's use and purpose and what appropriate data is likely to be available or collectable within the evaluation timescale.

In addition to impact, other key question areas to consider include:
- whether the needs of groups being targeted have been met (appropriateness and equity)

- how impact was possible or not in a particular setting that informs continuous improvement, scale and spread (process), and what costs have been avoided (efficiency)

- whether there is capacity for impact to sustain (sustainability)

Key Points
- Establish key evaluation questions based on the criteria considered to be a priority and relevant for the intended use of the evaluation (appropriateness, equity, process, impact, efficiency and sustainability)

Example set of evaluation questions
1. To what extent are medium to long term outcomes improving as a result of the programme? **(impact)**
2. To what extent has capacity and capability been developed that will ensure sustainability of improved outcomes in the medium to long term? **(impact and sustainability)**
3. To what extent is the programme engaging staff and users in planned activities and how well are these activities working (such as training, coaching etc.)? **(process/short term outcomes)**
4. What success factors can be identified as explaining how the programme is working in a particular setting? **(process)**
5. What range of outcomes (intended and unintended) has the programme contributed to in the medium to long term and how is this meeting the needs of different groups? **(impact, equity, appropriateness)**
6. To what extent have costs been avoided as a result of the programme? **(efficiency)**

What tools and resources can help?

Better Evaluation describe how key evaluation questions are developed and outline a range of other resources that can be used.

Step 4: Designing: developing the methods that will be used to answer the evaluation questions

The evaluation design determines how information will be collected and analysed to answer the evaluation questions. The exact design will depend on any constraints on accessing and analyzing existing data or collecting new data. The goal is to be able to gather the most reliable information about the programme within the time and resource constraints that will answer the key evaluation questions.

For an improvement programme a mixed design is likely to be used since there is often a focus on different evaluation criteria at the same time such as impact and process. As well as being theory-based to at least some extent, one of the following designs or a combination of these can be used:

Key Points

- The evaluation design will depend on available data, ability to get new data and level of resource

- There are a number of approaches to evaluation that can inform your design – each with different benefits depending on what you are looking to find out

Simple impact

Simple impact evaluation design focuses on understanding and describing whether there has been improvement or impact by comparing the programme to itself over time. This would be using a before and after design or time series analysis from a baseline.

Causal impact

Causal impact evaluation design makes a comparison with what was observed as a result of the programme with an estimate of what would have happened using a control group. For more guidance about causal impact evaluation see the case study based on an evaluation conducted by the EEvIT.

Process and case based

Process evaluation design is used to understand whether a programme was implemented as planned and how the process of change resulted in improvement or not and in what circumstances. Related to a process design is the use of a case-based approach to describe and/or compare particular instances of change or improvement as part of a programme. This is particularly useful when change is taking place across multiple sites or teams.

What tools and resources can help?

There are a range of resources available from Better Evaluation on the [design of theory-based and impact evaluations](#).

Step 5: Collecting and analysing: how information will be retrieved and used

A data collection approach will depend on what information will be necessary to describe and compare to satisfy the requirements of the evaluation design – simple or causal and/or process. Selecting appropriate measures or indicators of what is expected to improve over time (outcomes) and through what process will be the starting point. The following are key principles to consider:

- when to start collecting new data or retrieving existing data including having a suitable baseline and when specific outcomes are expected to be measurable – formative or real-time evaluation will mean collecting information as early as possible

- the frequency of collection required such as before and after vs repeated measurement over time (time series) and any sampling approach being used

- how qualitative narrative information will be collected to answer how and why questions that will include different perspectives
- whether causal impact is being assessed which would require comparison with a matched control group

Key Points
- Data collection depends on the design being used including the frequency of collection and level of comparison being used (simple vs causal)
- Think about how to align data collection and analysis with how components of the programme are expected to be delivered

An brief example of what collection and analysis would focus on for a collaborative learning programme:

- in the **short term** – how well participating staff have been engaged and responded to programme activities and what immediate benefit this is having,
- in the **medium term** – whether targeted practice is improving through tests of change at different sites,
- in the **long term** – on understanding the impact/overall difference such as organizational capacity and capability and quality of care (proxy measures of quality may be used such as emergency hospital admissions).

The availability of data is an important consideration for selecting measures/indicators – is there routine data already being collected such as the number of hospital admissions? A data sharing agreement may be required to ensure access to routinely collected data held elsewhere.

Step 5: Collecting and analysing: how information will be retrieved and used

What tools and resources can help?

- [Measurement tools](#) can be used to prioritise what information will be collected and how according to best practice including data collection checklist.

Step 5: Collecting and analysing: specific approaches to collecting data

There are a wide variety of existing tools available for use to support collection and analysis. Comparing pre-defined criteria on a scale through a questionnaire or survey tool is a common before and after measurement approach. Validated tools are recommended but it may also be important to tailor any tool to the local context.

It is unlikely that the use of predefined criteria in this way will provide the balance of detail and depth required to fully answer key evaluation questions. Narrative qualitative data is critical for understanding differential and holistic perspectives of whether and how impact has occurred.

Key Points

- Data collection depends on the design being used including the frequency of collection and level of comparison being used (simple vs causal)

- Think about how to align data collection and analysis with how the programme is being implemented and when and where outcomes are expected to occur

Using two or more methods to be able to triangulate the data enhances the credibility of the findings and the interpretations that can be made.

Case studies can be useful way of capturing learning and impact across a common unit such as organisation/service/team.

- Free text responses are a useful way of easily collecting perspective or views from people. Whereas, in-depth/semi-structured interviews or focus groups can be used to explore in detail the process and impact of a programme from those delivering and receiving the improvement of services.
- Observation provides a flexible way of assessing a process or situation that is under change by documenting what is seen and heard.
- Videoed or written patient (not all patients) stories can be used to prompt reflection and discussion as part of a formative evaluation approach or for capturing impact from the perspective of patients.
- Reflective information capture as part of a learning log or regular team discussion can be a flexible way of evaluating a programme during implementation. See here for an example.

Step 5: Collecting and analysing: specific approaches to collecting data

What tools and resources can help?

Guidance on how to [measure patient experience](#) from the Health Foundation is a useful starting point for exploring different approaches to capturing narrative from people receiving services.

Step 6: Synthesising and reporting: bringing information together to answer the evaluation questions

Synthesising and reporting what evaluation finds should be timed according to the how evaluation is being used. There may be different audiences being reached and different components of the evaluation that will be meaningful to share. Synthesising and reporting could be at regular interview during implementation (formative/real-time) or at the end of the programme in terms of its continuation and spread (summative).

Synthesising and making final interpretations about impact and learning involves making comparisons of change over time, for different measures or indicators, relevant to different groups and across different local contexts. There are specific considerations in terms of how data would be checked for validity, interpreted and displayed and produce a consolidated account of the evaluation findings.

Key Points
- Think about how different sources of data would be best brought together in a valid and clear way to answer the evaluation questions – could this involve comparison of data for different cases or key measures visualized over time as a whole for the programme?
- Communicating the findings of evaluation can involve reaching different audiences – how could visual summary and virtual dialogue be used to communicate key messages?

Qualitative narrative information can tell the story of impact and what happened during the programme to influence this. When compared across different individuals/teams/organisations this information can be a useful approach for identifying common lessons such as enablers or barriers.

How evaluation findings are shared should be tailored to the intended use of the evaluation (identified as part of the evaluation plan or framework). This could be in the form of a written report bringing together measurement data over time with explanatory qualitative data but could also include other forms: individual case studies and/or patient stories to communicate the findings for a particular setting or team
- visual summaries of findings including lessons learned
- workshops or virtual learning events

What tools and resources can help?

Detailed guidance about how to ensure a good standard of reporting specific to quality improvement can be found in the [Standards for Quality Improvement Reporting Excellence](#) (SQUIRE).

Evaluation planning list

Start to identify what type of evaluation is appropriate and feasible using evaluability	• Consider the purpose of evaluation and to what extent evaluability should be assed in order to clarify what approach would be feasible, credible and usable.
Involve stakeholders and end users	• Identify stakeholders and those that have an interest in the evaluation and plan how they will be supported to be involved in the evaluation at different stages. Their engagement supports the evaluation throughout the entire process.
Establish management and decision making processes	• Clarify how decisions will be made – should an evaluation steering or reference group be established
Clarifying who will conduct the evaluation and the resources required	• Skills and expertise of people internal and external to an organisation may be required and there should be a clear roles and responsibilities developed. • Depending on the design and methods, both internal resources (e.g. staff time) and external resources (e.g. participants' time to attend meetings to provide feedback) should be considered.
Document management processes and agreements	• Develop any formal documents needed, such as Terms of Reference.

Developing a framework for evaluation

An evaluation framework is a written document that describes the overall approach or strategy that will structure and guide the evaluation including how the programme theory is being defined that the evaluation will focus on. It includes the scope and purpose of the evaluation being conducted, what the evaluation will focus on including key evaluation questions that relate to the programme theory and how this will be answered using specific sources of information and data.

Table 1 summarises the evaluation framework developed for a national programme supporting the rollout of Near Me video consultation. The sources of existing and new data gathered and interpreted to answer the key evaluation questions are outlined where a focus on organization learning was prioritized along with programme evaluation. Measures or indicators of process and impact would be specified as the measurement or data collection plan is developed.
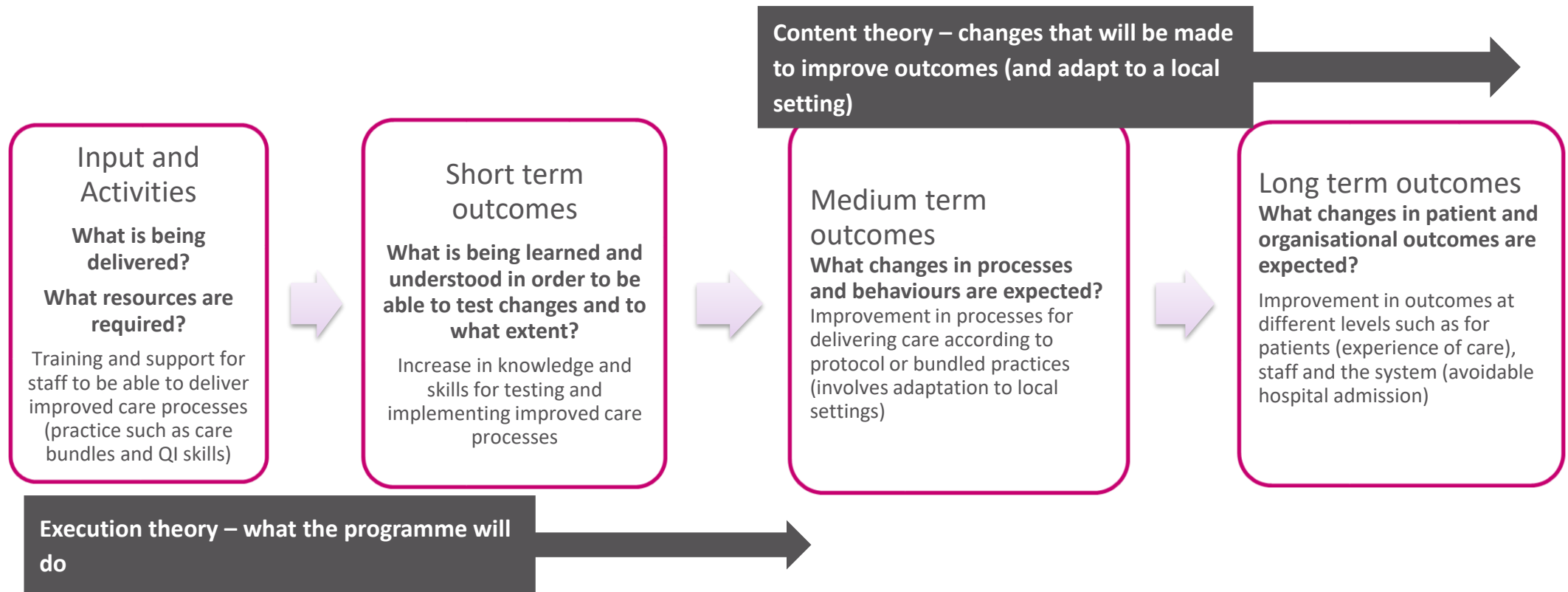
Table 1. Evaluation framework – key questions and how these will be answered

Evaluation focus	What questions are being addressed?	How will these be answered using existing and new data?
Organisational learning	• What are teams noticing and experiencing in terms of what has been successful and what has been more challenging? • What are teams learning from having worked in a new way to support improvement and that should inform future improvement work?	• Reflective qualitative data captured using a combination of semi-structured interviews, structured interviews and open ended survey questions • Documentary analysis • Shadowing of implementation
Programme impact evaluation	• To what extent is there improvement in Near Me use and in what circumstances? • What is being learned about the spread of Near Me use in practice including success and barriers?	• Measures/indicators of impact relating to the spread of Near Me use including increase in Near Me calls from baseline • Documentary analysis • Shadowing of implementation

Defining programme theory and outcomes

Defining when (and where/among who) outcomes are expected to occur and how these would be measured is important when evaluating the impact of quality improvement. The logic model below illustrates how this sequence of outcomes can be articulated.

Outcomes or impact at an organisational or system level commonly relate to what has improved in terms of efficiency and quality indicators such as service utilisation and satisfaction with care. At a personal outcome level, impact relates to what matters to service users to be able to live well in the context of their lives.

Content theory – changes that will be made to improve outcomes (and adapt to a local setting)

Input and Activities

What is being delivered?

What resources are required?

Training and support for staff to be able to deliver improved care processes (practice such as care bundles and QI skills)

Short term outcomes

What is being learned and understood in order to be able to test changes and to what extent?

Increase in knowledge and skills for testing and implementing improved care processes

Medium term outcomes

What changes in processes and behaviours are expected?
Improvement in processes for delivering care according to protocol or bundled practices (involves adaptation to local settings)

Long term outcomes

What changes in patient and organisational outcomes are expected?

Improvement in outcomes at different levels such as for patients (experience of care), staff and the system (avoidable hospital admission)

**Execution theory – what the programme will do**

Using wider theory for evaluation

There are different levels of theory that can be used to develop a programme's theory of change and how an evaluation should be structured. The Kirkpatrick model is a widely used for articulating the outcomes expected from training across four levels: reaction, learning, behavior, and results. A limitation of the Kirkpatrick model is that it does not include other factors that will influence improvement in knowledge such as individual human factors (motivation) and organizational factors (culture).

Programme theory should also take account of the factors that are not within the direct control of the programme but may nevertheless influence whether there is improvement. For instance, factors in relation to organizational culture may influence the extent to which participants are able to engage with the programme and put learning into practice. The use of broader frameworks such as the Model for understanding Success in Quality can be used to clarify what other factors would be appropriate for an evaluation to assess.

There can also be factors that relate to the individual characteristics of participants such as the level of motivation to learn or prior experience of quality improvement. These factors are sometimes referred to as moderators. There might be ways that the programme can seek to control these factors as part of the programme design. Such as how selection of those participating in the programme can be based on evidence of their prior experience and demonstration of motivation.

Level 4: Organisational improvement

Level 3: Behavioural change

Level 2: Learning

Level 1: Reaction
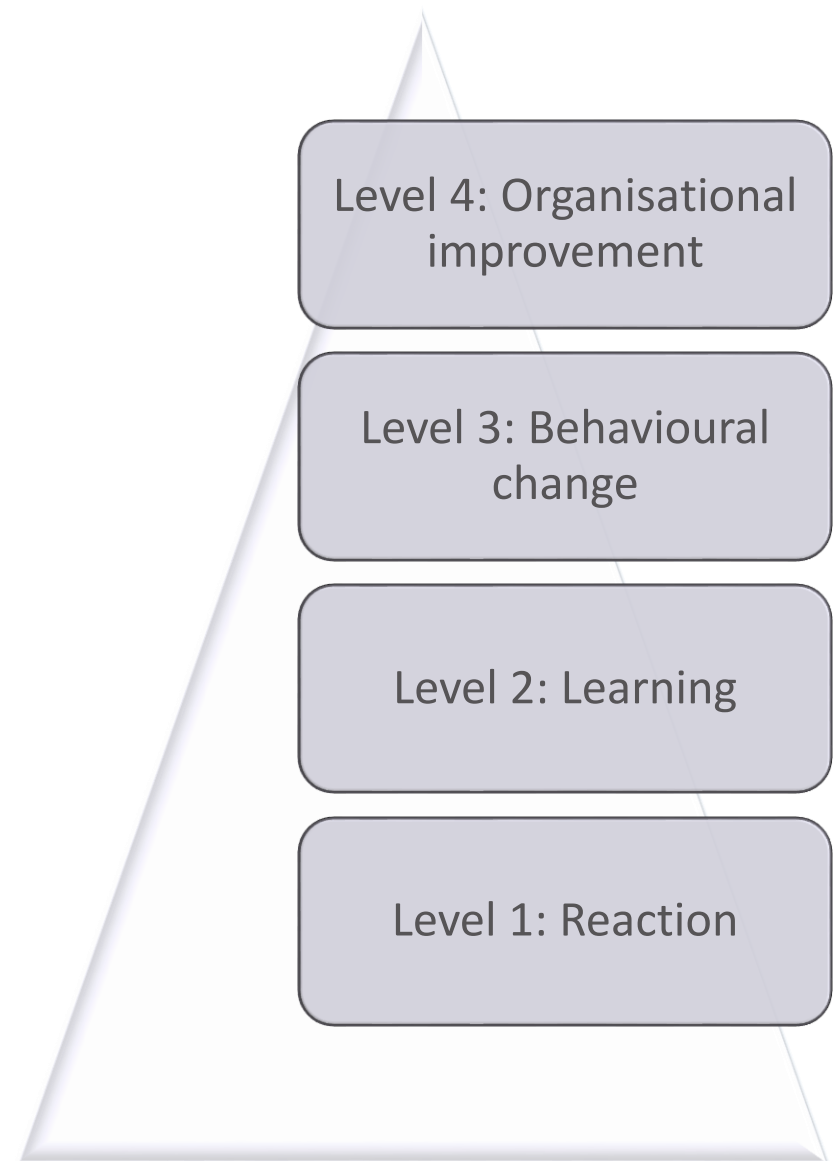
Impact evaluation example

Changes in emergency admission, 28-day re-admission and length of stay following introduction of a new model of care provided a focus for measuring the expected impact of a programme evaluated here. Time series analysis illustrates that for one practice implementing the model there had been a sustained downwards shift in the rate of emergency hospital admissions for patients over 65 years.

These results have to be interpreted with caution due to the issue of 'regression to the mean'. This relates to how the health status or hospital use of patients can improve over time on average as people are expected to get better, especially if patients being referred to a new service or model had a greater need for healthcare utilization at the start (recent crisis or emergency hospital admission).

Chart 1: Practice 1 - emergency admissions age 65+

In the context of evaluating health care initiatives, the health status or hospital use of patients can improve over time as on average people would be expected to get better over time. A more robust way of assessing impact when this is expected is to compare with a matched control group. A group of patients would be selected to be as similar as possible (such as in terms of age, gender, prior health conditions, access to health care services and prior use of hospital services).
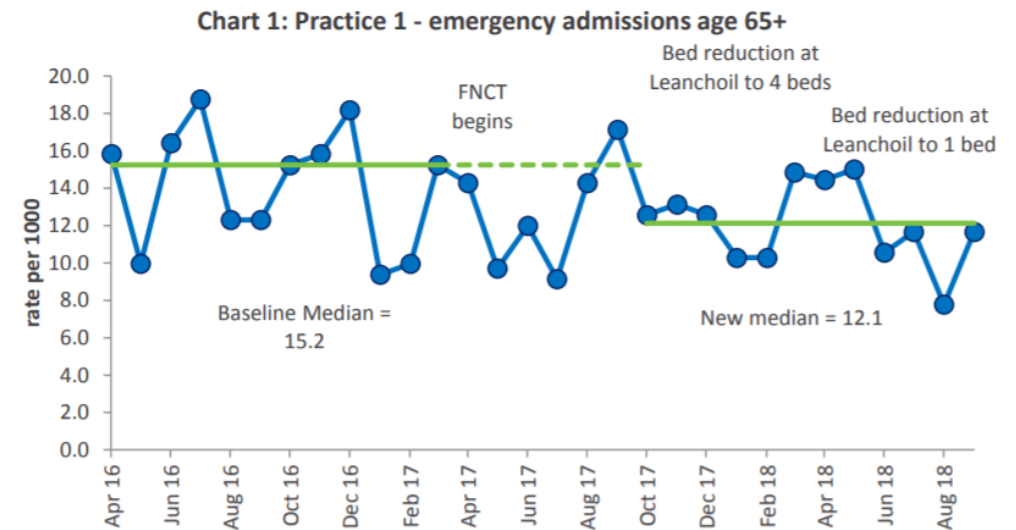
Reflective questions

Reflective questions can be used to collect information about progress, explore the process of emergent change and identify learning. Reflections can be from an individual perspective collected through a reflective log or survey. They can also be from a team perspective prompted through team huddles or discussion.

What has happened?

How are you feeling?

What are you noticing that is working well?

What are you noticing this is working less well? How should this be different?